



AI agents in radiology: toward autonomous and adaptive intelligence

Burak Koçak¹
 İsmail Meşe²

¹Başakşehir Çam and Sakura City Hospital,
Department of Radiology, İstanbul, Türkiye

²Üsküdar State Hospital, Clinic of Radiology, İstanbul,
Türkiye

The integration of artificial intelligence (AI) into healthcare is no longer a futuristic concept but a present-day reality, with radiology at the forefront of this technological wave. Although generative AI applications are now well established, a new paradigm is gaining prominence. This commentary focuses on advanced AI systems often broadly referred to as AI agents, specifically those exhibiting highly agentic capabilities, which are beginning to be distinguished as agentic AI.¹ An AI agent can be defined as a sophisticated system capable of autonomous, goal-directed reasoning, integrating planning, memory, tool usage, and feedback.

AI agents promise a higher degree of autonomy and adaptability, capable of independent operation, learning, and even collaboration within complex clinical settings, potentially with reduced need for direct human oversight.² Although the theoretical and practical boundaries of AI agents and agentic AI remain under exploration in academic discourse, early evidence indicates their capacity to redefine healthcare delivery.³⁻⁵ For radiology, this could signify a shift toward an agentic era, where such intelligent systems are embedded across the imaging workflow, automating protocol selection, interpreting studies, generating reports, and interacting with radiologists to support complex diagnostic decisions.⁵ This commentary with brief literature review explores AI agents within the radiological context, outlines their evolutionary path from prior AI systems, and explores their potential applications and inherent challenges, offering a conceptual overview rather than an exhaustive technical review.

Distinguishing artificial intelligence agents

Understanding the leap to highly agentic AI systems requires examining the recent evolution of AI (Table 1 and Figure 1), particularly systems built on large language models (LLMs). Initially, based on their training data, foundational LLMs offered impressive text generation and comprehension. The next iteration saw these models augmented, equipped with tools such as retrieval augmented generation or connections to external databases and software, allowing them to access current information or perform specific tasks beyond their core training.⁶ However, these prior approaches often still rely on human direction or supervision for complex operations.

AI agents, particularly those exhibiting the agentic characteristics central to this discussion, represent the next step.^{2,5,7} An AI agent is not just a model; it is a system orchestrated by a core reasoning engine (often an LLM) that can autonomously decompose a high-level goal into smaller, executable steps. It can select and use different tools (e.g., segmentation algorithms, data retrieval functions), store key information in its memory, and adapt its plan based on intermediate results. This ability to self-direct and adapt significantly distinguishes them from their predecessors.

To illustrate this with a neuroradiological example, the workup for a patient presenting with acute stroke symptoms can be considered. An early vision-language model (VLM) might be given a single image and asked to describe it [e.g., “non-contrast computed tomography (CT) shows a hyperdense left middle cerebral artery (MCA) sign”]. An augmented LLM, when prompted by a radiologist, could retrieve the text from the non-contrast CT and CT angiography reports to help draft a summary. In contrast, an AI agent, given the high-level goal “Evaluate for acute stroke for patient X,” could autonomously execute a complex and time-sensitive

Corresponding author: Burak Koçak

E-mail: drburakkocak@gmail.com

Received 20 May 2025; revision requested 10 June 2025;
accepted 25 June 2025.



Epub: 07.07.2025

Publication date: xx.xx.2025

DOI: 10.4274/dir.2025.253470

plan as follows: (i) access the initial non-contrast head CT and activate a tool to calculate an Alberta Stroke Program Early CT score; (ii) open the subsequent CT angiogram to

identify the vessel occlusion location; (iii) run a perfusion imaging tool to calculate the ischemic core and penumbra volumes; (iv) synthesize these quantitative findings with clinical guidelines; and (v) generate a preliminary report summarizing that the patient has a left M1 occlusion with a small core and large penumbra, flagging the case for immediate review as a potential thrombectomy candidate. This example demonstrates a transition from a reactive tool to a proactive, autonomous assistant orchestrating a critical and highly complex diagnostic workflow.

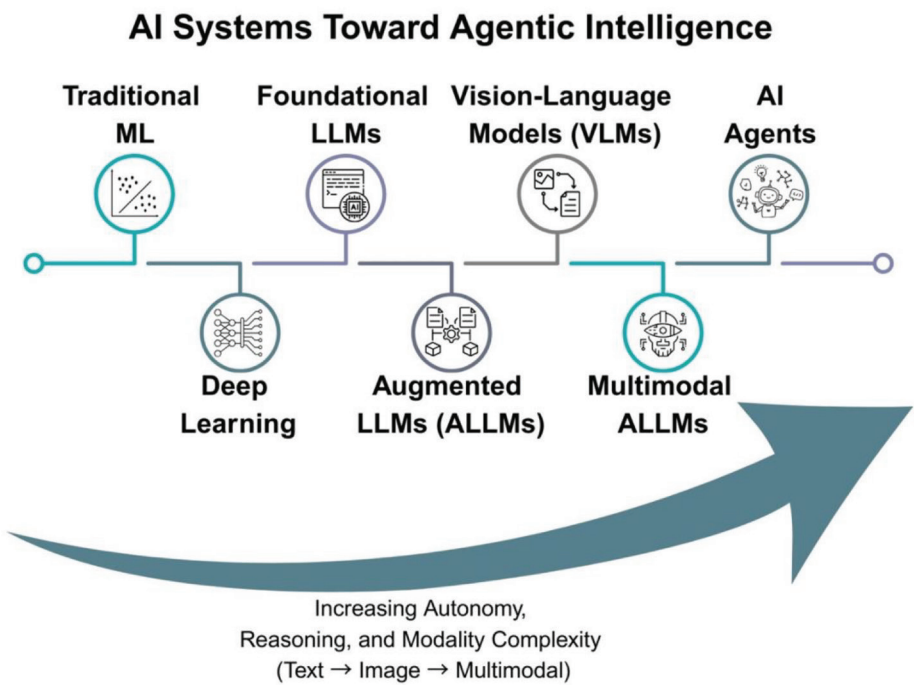


Figure 1. Oversimplified conceptual evolution of artificial intelligence (AI) from traditional machine learning models to autonomous AI agents. The placement of vision-language models and augmented large language models is illustrative, and their development may overlap. ML, machine learning; LLM, large language model; VLM, vision-language model; ALLM, augmented large language model.

Potential applications in radiology

The advent of AI agents heralds a significant transformation in radiological practice, offering pathways to optimize complex clinical workflows and enhance diagnostic capabilities.^{5,8} Figure 2 presents a simplified AI agent-based workflow example in radiology, demonstrating how such agents could orchestrate various clinical tasks while leaving final decision-making in the hands of radiologists. Below, we outline a few key areas where AI agents may be integrated into radiology practice.

Table 1. Conceptual progression of artificial intelligence systems toward agentic intelligence			
Type	Key characteristics	Key limitations	Potential radiological capabilities
Traditional ML	Algorithm-based (e.g., SVM); relies on structured data and handcrafted feature engineering	Limited scalability for complex data, which are less adaptable; often requires domain expertise for feature selection	Basic image analysis; statistical analysis of patient data; risk prediction from structured data
Deep learning	Neural network-based; learns features automatically from large datasets (images, text, etc.)	Data-hungry; computationally intensive; can be a “black box” (interpretability challenges)	Advanced image segmentation, classification, anomaly detection in medical images (e.g., lesion detection)
Foundational LLMs	Pretrained on large-scale text data; operate on fixed knowledge; generate coherent text	Primarily text-based; no external tool use; no real-time data access; lack dynamic planning or memory	Summarizing findings from text reports; assisting with unstructured textual reports; drafting initial report sections
Augmented LLMs	Foundational (text-based) LLMs enhanced with tool use (e.g., RAG, calculators, APIs) via orchestration or plug-ins	Still primarily text-centric if not combined with a VLM; autonomy limited by tool capabilities and prompting	Retrieving priors (textual medical records); assisting protocoling (via text-based queries); citing literature; using calculators for quantitative data
VLMs	Pretrained on large-scale image-text data; understand and generate text related to visual inputs; multimodal outputs	Operate on fixed knowledge (unless augmented); may lack sophisticated reasoning beyond pattern matching without further integration	Basic image captioning; visual question answering (about images); identifying findings in images based on textual prompts
Multimodal ALLMs	Integrates VLM capabilities with ALLM tool use; processes and reasons over multiple modalities (text, image, etc.) using external tools/data	Increased complexity in orchestration; potential for compounded errors from different modules; resource-intensive	Interpreting medical images using vision tools combined with RAG; cross-modal report generation; advanced visual question answering with external knowledge
Artificial intelligence agents	Autonomous systems integrating LLMs/VLMs/ALLMs, tool use, memory, planning, and reinforcement learning or feedback	Complex to govern; opaque reasoning; require robust oversight for safety, ethical considerations, and accountability	Automated triage; protocol selection; complex segmentation and detection; structured report generation; dynamic question answering; multimodal synthesis; decision support

ML, machine learning; SVM, support vector machine; VLM, vision-language model; RAG, retrieval-augmented generation; API, application programming interface; LLM, large language model; ALLM, augmented large language model.

AI Agent–Enabled Radiology Workflow

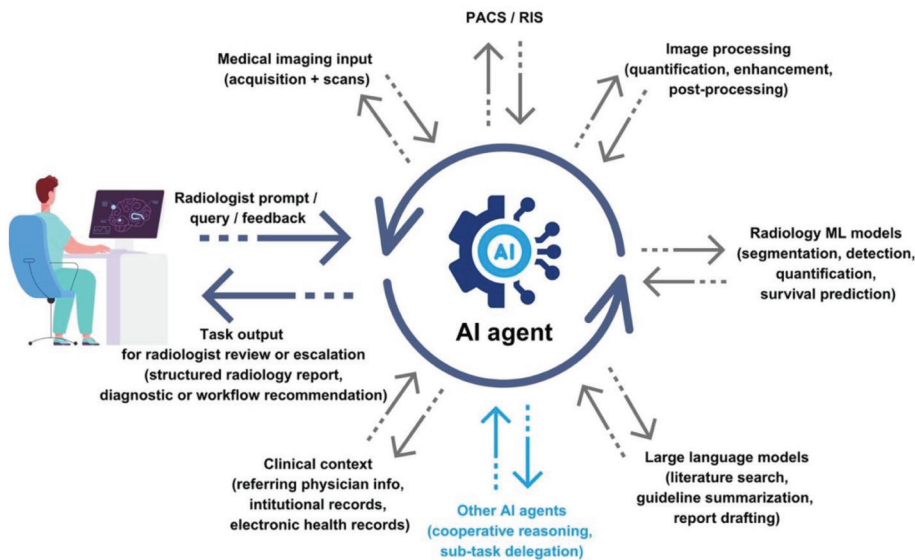


Figure 2. Example of an oversimplified artificial intelligence (AI) agent-enabled radiology workflow, where the agent acts as a background orchestrator, performing several tasks and escalating ambiguous cases for the radiologist's review. PACS, picture archiving and communication system; ML, machine learning; RIS, radiology information system.

Automation of administrative and preparatory tasks

One of the most immediate impacts could be the automation of laborious preparatory and administrative duties.⁵ Imagine intelligent systems that efficiently triage imaging studies based on urgency, recommend optimal imaging protocols, or collate pertinent patient histories from disparate electronic health records. Such automation would liberate radiologists from these routine tasks, allowing them to channel their cognitive expertise toward more intricate image analysis and critical diagnostic decision-making.

Image analysis and structured reporting

More sophisticated AI agents, especially when embedded within existing radiology platforms, could further amplify their value by concurrently analyzing imaging data, contextualizing findings against current medical literature, and even drafting preliminary structured reports, thereby cultivating a more efficient and accurate diagnostic pipeline.³

An illustrative advancement in this domain is RadGPT, a specialized vision-language AI agent designed for generating comprehensive reports from abdominal CT scans.⁹ This system reportedly not only segments tumors and adjacent anatomical structures but also produces both structured and narrative summaries, detailing characteristics such as

tumor dimensions, morphology, location, attenuation, volume, and its relationship with nearby vasculature and organs. The system's reported high sensitivity and specificity, particularly for detecting small tumors, underscores its potential. The RadGPT system employs deterministic algorithms to translate voxel-level annotations into structured data, which are then processed by LLMs to create narrative reports, potentially enriching radiologists' reports with precise details, such as tumor volume and attenuation, that might otherwise be overlooked.

Multimodal integration for diagnostic support

Synergy between multimodal LLMs and AI agents can be harnessed to integrate diverse radiological and clinical data streams for enhanced diagnostic support.⁵ These agents could interface with picture archiving and communication systems (PACS) to automate quality assurance processes, manage data transfers, execute image analysis algorithms, and flag potential abnormalities for radiologist review. Operating in the background, such agents can continuously process imaging data, generate initial findings, and propose differential diagnoses. As complementary or embedded tools (i.e., multimodal LLM + agents), VLMs further empower radiologists by facilitating structured report generation, augmenting the review process, enabling visual search capabilities,

and summarizing extensive patient imaging histories. For example, systems such as LLaVA-Med, specifically trained on biomedical datasets, have demonstrated proficiency in image interpretation, clinical reporting, and responding to visual queries.¹⁰

Dynamic task execution with external tools

AI agents that can dynamically plan and execute tasks using external tools show promise. For example, VoxelPrompt has reportedly surpassed task-specific models in complex tasks such as image segmentation and pathology characterization.¹¹ VoxelPrompt functions as an agent-driven vision-language framework. It receives a natural language prompt and three-dimensional medical volumes, and its core LLM-based controller (or agent) iteratively predicts the executable instructions. These instructions are not simple commands; they can involve interacting with dedicated vision networks, calling a predefined library of functions, and interpreting intermediate results.

Potential challenges and risks

Despite the considerable promise of AI agents in healthcare, their widespread adoption, particularly in a critical field such as radiology, is contingent upon addressing significant inherent challenges.^{2,5}

The development and enforcement of robust governance structures, alongside the evolution of legal and ethical guidelines, are paramount to nurturing innovation while safeguarding against potential pitfalls related to algorithmic bias, accountability, and the establishment of trust.² AI agents can inadvertently amplify underlying model biases, introducing safety risks.^{12,13} Addressing these risks effectively remains a formidable task in the healthcare domain, especially as regulatory frameworks often struggle to keep pace with rapid technological advancements. Moreover, the continuous learning capabilities of some agentic AI systems further complicate regulatory oversight, as their evolving behavior can challenge static approval processes, necessitating more dynamic and adaptive regulatory strategies.

Transparency in AI agent operations is crucial for gaining clinician trust and facilitating smoother implementation, particularly given the understandable reluctance to adopt black-box technologies.¹⁴ Therefore, designing AI agents with an emphasis on explainability, mechanisms to address biases, and robust security in their decision-making pathways is essential, especially in high-

stake and trust-sensitive environments such as medical diagnostics.

Integration of highly autonomous agents introduces significant human–AI interaction risks.^{15,16} Automation bias, the tendency for humans to over-accept suggestions from automated systems, could lead to missed diagnoses if radiologists become less vigilant. Similarly, over-reliance on AI-generated outputs might deskill practitioners over time, diminishing their ability to interpret complex cases without AI assistance. Mitigating these risks requires not only robust AI validation but also targeted training for clinicians on the appropriate use and limitations of AI agents, leading to a culture of critical engagement rather than passive acceptance.

The practical deployment of AI agents in clinical settings also presents certain hurdles, particularly concerning security and patient privacy.¹⁷ These agents are often envisioned to require access to sensitive patient data and possess the capability to execute actions autonomously. This, coupled with their reliance on natural language communication, can introduce new security vulnerabilities. Ensuring secure memory management is vital to counter potential threats, including the risk of reintroducing corrupted or poisoned data during information retrieval processes.⁷ Implementing rigorous auditing of tool usage has been suggested to prevent unauthorized actions and data breaches, although this may result in substantial computational costs.^{7,18} There is also a need for universally accepted safety evaluation benchmarks and consensus on design standards across the AI agent ecosystem.

From an operational standpoint, integrating these technological innovations into established daily clinical workflows remains a significant barrier.⁵ Technical difficulties include a dependency on high-quality, comprehensively labeled datasets, which can be particularly scarce for specialized or rare medical conditions.⁶ Furthermore, system integration presents complexities, as AI agents must interoperate flawlessly with existing heterogeneous hospital infrastructures, such as PACS and electronic health records, each often possessing distinct standards in terms of data and management.³ Finally, the computational resources required to deploy and maintain powerful deep learning systems for real-time performance at scale, especially in resource-constrained remote or edge computing scenarios, can pose a significant limitation.¹⁹

In conclusion, AI agents represent a significant evolutionary step in AI, holding substantial potential to elevate diagnostic and decision support, streamline workflow efficiency, and ultimately improve patient care in radiology. Their inherent capacity for autonomous, goal-oriented behavior, coupled with their ability to synthesize diverse data types and utilize external tools, sets them apart from the earlier generations of AI systems. However, the journey toward their widespread and safe implementation is paved with challenges, particularly concerning transparency, algorithmic bias, human–AI interaction risks (e.g., automation bias), data security, and system interoperability. Effectively navigating these obstacles will necessitate robust governance frameworks, collaborative interdisciplinary research, and flexible and adaptive regulatory approaches. Although the field of AI agents in radiology is still in its nascent stages, their integration into radiological practice seems inevitable, with profound implications for both the delivery of clinical care and the optimization of operational processes in the years to come.

Acknowledgement

Language of this manuscript was checked and improved by ChatGPT (4o) and Gemini 2.5 Pro Preview. The authors conducted strict supervision when using these tools.

Footnotes

Conflict of interest disclosure

Burak Koçak, MD, serves as Section Editor for Diagnostic and Interventional Radiology. He had no involvement in the peer review of this article and had no access to information regarding its peer review.

References

1. Sapkota R, Roumeliotis KI, Karkee M. AI agents vs. agentic AI: a conceptual taxonomy, applications and challenges. *arXiv*. 2025. [\[Crossref\]](#)
2. Hughes L, Dwivedi YK, Malik T, et al. AI agents and agentic systems: a multi-expert analysis. *Journal of Computer Information Systems*. 2025;0(0):1-29. [\[Crossref\]](#)
3. Patel D, Raut G, Cheetirala SN, et al. AI agents in modern healthcare: from foundation to pioneer -- a comprehensive review and implementation roadmap for impact and integration in clinical settings. Preprints. 2025. [\[Crossref\]](#)
4. Hinostroza Fuentes VG, Karim HA, Tan MJT, Aïdahloul N. AI with agency: a vision for adaptive, efficient, and ethical healthcare. *Front Digit Health*. 2025;7:1600216. [\[Crossref\]](#)

5. Karunanayake N. Next-generation agentic AI for transforming healthcare. *Informatics and Health*. 2025;2(2):73-83. [\[Crossref\]](#)
6. Mialon G, Dessi R, Lomeli M, et al. Augmented language models: a survey. *arXiv*. 2023. [\[Crossref\]](#)
7. Deng Z, Guo Y, Han C, et al. AI agents under threat: a survey of key security challenges and future pathways. *ACM Comput Surv*. 2025;57(7):1-36. [\[Crossref\]](#)
8. Khosravi B, Rouzrokh P, Faghani S. AI agents in radiology: the future of intelligent workflows. Last Accessed date: 16.05.2025. [\[Crossref\]](#)
9. Bassi PRAS, Yavuz MC, Wang K, et al. RadGPT: constructing 3D image-text tumor datasets. *arXiv*. 2025. [\[Crossref\]](#)
10. Li C, Wong C, Zhang S, et al. LLaVA-Med: training a large language-and-vision assistant for biomedicine in one day. *arXiv*. 2023. [\[Crossref\]](#)
11. Hoopes A, Butoi VI, Gutttag JV, Dalca AV. VoxelPrompt: a vision-language agent for grounded medical image analysis. *arXiv*. 2024. [\[Crossref\]](#)
12. Koçak B, Ponsiglione A, Stanzione A, et al. Bias in artificial intelligence for medical imaging: fundamentals, detection, avoidance, mitigation, challenges, ethics, and prospects. *Diagn Interv Radiol*. 2025;31(2):75-88. [\[Crossref\]](#)
13. Tian Y, Yang X, Zhang J, Dong Y, Su H. Evil geniuses: delving into the safety of LLM-based agents. *arXiv*. 2024. [\[Crossref\]](#)
14. Zhang J, Zhang Zm. Ethics and governance of trustworthy medical artificial intelligence. *BMC Med Inform Decis Mak*. 2023;23(1):7. [\[Crossref\]](#)
15. Park SH, Langlotz CP. Crucial role of understanding in human-artificial intelligence interaction for successful clinical adoption. *Korean J Radiol*. 2025;26(4):287-290. [\[Crossref\]](#)
16. Dratsch T, Chen X, Rezazade Mehrizi M, et al. automation bias in mammography: the impact of artificial intelligence BI-RADS suggestions on reader performance. *Radiology*. 2023;307(4):e222176. [\[Crossref\]](#)
17. Akinci D'Antonoli T, Tejani AS, Khosravi B, et al. Cybersecurity threats and mitigation strategies for large language models in health care. *Radiol Artif Intell*. 2025:e240739. [\[Crossref\]](#)
18. Zhang X, Xu H, Ba Z, Wang Z, Hong Y, Liu J. PrivacyAsst: safeguarding user privacy in tool-using large language model agents. *IEEE Transactions on Dependable and Secure Computing*. 2024;21(6):5242-5258. [\[Crossref\]](#)
19. Kocak B, Ponsiglione A, Romeo V, Ugga L, Huisman M, Cuocolo R. Radiology AI and sustainability paradox: environmental, economic, and social dimensions. *Insights Imaging*. 2025;16(1):88. [\[Crossref\]](#)